

R version 4.2.2 (2022-10-31 ucrt) -- "Innocent and Trusting"
 Copyright (C) 2022 The R Foundation for Statistical Computing
 Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
 You are welcome to redistribute it under certain conditions.
 Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
 Type 'contributors()' for more information and
 'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
 'help.start()' for an HTML browser interface to help.
 Type 'q()' to quit R.

[Previously saved workspace restored]

```
> Cancer_data=read.csv(C:/Users/yxa210024/Desktop/Masters/spring2023/Stats for DS/mini_project6/prostate_cancer.csv')
Error: unexpected '/' in "Cancer_data=read.csv(C:/"
> Cancer_data=read.csv(C:/Users/yxa210024/Desktop/Masters/spring2023/Stats for DS/mini_project6/prostate_cancer.csv')
Error: unexpected 'for' in "Cancer_data=read.csv(C:/Users/yxa210024/Desktop/Masters/spring2023/Stats for"
> Cancer_data=read.csv(C: Users/yxa210024/Desktop/Masters/spring2023/Stats for DS/mini_project6/prostate_cancer.csv')
Error: unexpected 'for' in "Cancer_data=read.csv(C: Users/yxa210024/Desktop/Masters/spring2023/Stats for"
> Cancer_data=read.csv(C: Users/yxa210024/Desktop/Masters/spring2023/StatsforDS/mini_project6/prostate_cancer.csv')
+
Error: unexpected end of input
> Cancer_data=read.csv(C: /Users/yxa210024/Desktop/Masters/spring2023/StatsforDS/mini_project6/prostate_cancer.csv')
Error: unexpected '/' in "Cancer_data=read.csv(C: /"
> Cancer_data=read.csv(C:/Users/yxa210024/Desktop/Masters/spring2023/StatsforDS/mini_project6/prostate_cancer.csv')
Error: unexpected '/' in "Cancer_data=read.csv(C:/"
> Cancer_data=read.csv("C:/Users/yxa210024/Desktop/Masters/spring2023/Stats for DS/mini_project6/prostate_cancer.csv")
Error: unexpected input in "Cancer_data=read.csv("
> Cancer_data=read.csv("C:/Users/yxa210024/Desktop/Masters/spring2023/Stats for DS/mini_project6/prostate_cancer.csv")
> attach(Cancer_data)
> plot(cancervol,log(psa))
> plot(cancervol,log(psa), main = "cancervol vs psa log")
Warning messages:
1: unable to open printer
2: opening device failed
> plot(cancervol,log(psa), main = "cancervol vs psa log")
> plot(cancervol,log(psa), main = "cancervol scatterplot")
> plot(cancervol,log(psa), main = "Cancervol Scatter plot")
> cor(cancervol,log(psa))
[1] 0.6570739
> plot(weight,log(psa), main = " Weight Scatter plot ")
> cor(weight,log(psa))
[1] 0.1217208
> plot(age,log(psa), main = " Age Scatter plot ")
> cor(age,log(psa))
[1] 0.1699068
> plot(benpros,log(psa), main = " Benpros Scatter plot ")
> cor(benpros,log(psa))
[1] 0.1574016
> plot(vesinv,log(psa), main = " Vesinv Scatter plot ")
> cor(vesinv,log(psa))
```

```

[1] 0.5663641
>
> plot(capspen,log(psa), main = " Capspen Scatter plot ")
> cor(capspen,log(psa))
[1] 0.5180231
>
> plot(gleason,log(psa), main = " Gleason Scatterplot ")
> cor(gleason,log(psa))
[1] 0.5390167
> plot(gleason,log(psa), main = " Gleason Scatter plot ")
> cor(gleason,log(psa))
[1] 0.5390167
> #None of the above variables show a linear trend with the psa-response variable
> cor(cancervol,psa)
[1] 0.6241506
> #we could not show a linear trend with psa-response variable from any above variables
> cor(cancervol,psa)
[1] 0.6241506
> # Cancervol and psa have a strong positive correlation
> table(cancer_data$vesinv)
Error in table(cancer_data$vesinv) : object 'cancer_data' not found
> table(cancer_data$gleason)
Error in table(cancer_data$gleason) : object 'cancer_data' not found
> table(Cancer_data$vesinv)

 0  1
76 21
> table(Cancer_data$gleason)

 6  7  8
33 43 21
> # individual variables
> Ind_varia_1 <- lm(log(psa)~weight+age+benpros)
> summary(Ind_varia_1)

Call:
lm(formula = log(psa) ~ weight + age + benpros)

Residuals:
    Min       1Q   Median       3Q      Max
-2.6950 -0.7076 -0.0243  0.6254  3.0399

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.065917   1.051430   1.014   0.313
weight       0.001837   0.002707   0.679   0.499
age          0.019484   0.016907   1.152   0.252
benpros      0.033462   0.043265   0.773   0.441

Residual standard error: 1.146 on 93 degrees of freedom
Multiple R-squared:  0.04406,    Adjusted R-squared:  0.01322
F-statistic: 1.429 on 3 and 93 DF,  p-value: 0.2394

> anova(Ind_varia_1)
Analysis of Variance Table

Response: log(psa)
      Df Sum Sq Mean Sq F value Pr(>F)
weight  1   1.893  1.89301   1.4414 0.2330
age      1   2.951  2.95084   2.2468 0.1373
benpros  1   0.786  0.78558   0.5982 0.4412
Residuals 93 122.139 1.31333
> step(nullmd,scope=list(lower=~1,upper=~weight+age+benpros))
Error in terms(object) : object 'nullmd' not found
> nullmd = lm(log(Cancer_data$psa)~1, data= Cancer_data)
> step(nullmd,scope=list(lower=~1,upper=~weight+age+benpros))
Start: AIC=28.72
log(Cancer_data$psa) ~ 1

```

	Df	Sum of Sq	RSS	AIC
+ age	1	3.6885	124.08	27.883
+ benpros	1	3.1655	124.60	28.291
<none>			127.77	28.725
+ weight	1	1.8930	125.88	29.277

Step: AIC=27.88

log(Cancer_data\$psa) ~ age

	Df	Sum of Sq	RSS	AIC
<none>			124.08	27.883
- age	1	3.6885	127.77	28.725
+ benpros	1	1.3363	122.74	28.833
+ weight	1	1.1554	122.92	28.976

Call:

lm(formula = log(Cancer_data\$psa) ~ age, data = Cancer_data)

Coefficients:

(Intercept)	age
0.79721	0.02633

> nullmd

Call:

lm(formula = log(Cancer_data\$psa) ~ 1, data = Cancer_data)

Coefficients:

(Intercept)
2.479

> step(nullmd,scope= list(lower=~1, upper=~cancervol + as.factor(vesinv) +
+ capspen +gleason), k = log(97))

Start: AIC=31.3

log(Cancer_data\$psa) ~ 1

	Df	Sum of Sq	RSS	AIC
+ cancervol	1	55.164	72.605	-18.9492
+ as.factor(vesinv)	1	40.984	86.785	-1.6449
+ gleason	1	37.122	90.647	2.5788
+ capspen	1	34.286	93.482	5.5663
<none>			127.769	31.2993

Step: AIC=-18.95

log(Cancer_data\$psa) ~ cancervol

	Df	Sum of Sq	RSS	AIC
+ gleason	1	8.247	64.358	-26.070
+ as.factor(vesinv)	1	6.547	66.058	-23.541
<none>			72.605	-18.949
+ capspen	1	0.967	71.638	-15.675
- cancervol	1	55.164	127.769	31.299

Step: AIC=-26.07

log(Cancer_data\$psa) ~ cancervol + gleason

	Df	Sum of Sq	RSS	AIC
+ as.factor(vesinv)	1	4.0178	60.340	-27.7480
<none>			64.358	-26.0697
+ capspen	1	0.1685	64.190	-21.7493
- gleason	1	8.2468	72.605	-18.9492
- cancervol	1	26.2887	90.647	2.5788

Step: AIC=-27.75

log(Cancer_data\$psa) ~ cancervol + gleason + as.factor(vesinv)

	Df	Sum of Sq	RSS	AIC
<none>			60.340	-27.748
- as.factor(vesinv)	1	4.0178	64.358	-26.070

```
+ capspen      1      0.3013 60.039 -23.659
- gleason      1      5.7179 66.058 -23.541
- cancervol    1     12.7041 73.044 -13.789
```

```
Call:
lm(formula = log(Cancer_data$psa) ~ cancervol + gleason + as.factor(vesinv),
    data = Cancer_data)
```

```
Coefficients:
```

```
(Intercept)      cancervol      gleason  as.factor(vesinv)1
   -0.72120      0.05981      0.38491      0.62117
```

```
> lm_model <- lm(formula = log(Cancer_data$psa) ~ cancervol + gleason + as.factor(vesinv),
+ data = Cancer_data)
> fitted(lm_model)
```

1	2	3	4	5	6	7	8
1.621772	1.995423	2.009114	1.606299	1.714907	1.609211	1.713645	1.707532
9	10	11	12	13	14	15	16
2.000615	1.662816	1.665083	1.603786	2.272428	2.235948	2.173779	1.867286
17	18	19	20	21	22	23	24
2.012494	2.178928	1.622448	2.044805	1.777184	2.442488	1.623136	2.327881
25	26	27	28	29	30	31	32
1.676626	1.843271	2.072803	2.013289	2.142419	2.254239	1.668217	1.659891
33	34	35	36	37	38	39	40
1.803406	1.648700	1.647498	2.194875	2.605562	2.067946	3.449465	2.106310
41	42	43	44	45	46	47	48
2.469297	2.225649	1.695109	2.324352	2.238592	2.287770	3.896381	2.163994
49	50	51	52	53	54	55	56
1.932482	1.790881	2.151098	1.902856	2.072803	2.476516	3.753908	2.186184
57	58	59	60	61	62	63	64
2.130977	1.683032	2.460750	2.145841	1.683032	3.036327	3.322233	3.049784
65	66	67	68	69	70	71	72
2.062287	2.230751	2.424084	2.513007	1.626420	2.169808	2.978583	2.163994
73	74	75	76	77	78	79	80
2.794946	3.355889	4.180648	3.591359	2.804512	3.116519	3.440954	3.322233
81	82	83	84	85	86	87	88
2.233335	2.709187	3.407764	3.230486	2.642748	4.600958	2.039170	2.931750
89	90	91	92	93	94	95	96
3.972762	3.263915	3.130873	3.345224	3.992835	5.706992	4.077249	4.044801
97							

```
4.901464
```

```
> resid(lm_model)
```

1	2	3	4	5	6
-2.051017742	-2.155592069	-2.169283217	-1.766467289	-1.344723308	-0.839103198
7	8	9	10	11	12
-0.943536330	-0.857380755	-0.950492987	-0.612693563	-0.395041143	-0.333744995
13	14	15	16	17	18
-1.002386909	-0.886058813	-0.773828390	-0.417312530	-0.542547954	-0.688949779
19	20	21	22	23	24
-0.062410428	-0.444811303	-0.137216562	-0.782547145	0.076873774	-0.617873784
25	26	27	28	29	30
0.053434939	-0.073245855	-0.272744852	-0.193266463	-0.292390854	-0.364294235
31	32	33	34	35	36
0.251789351	0.350066681	0.206551871	0.371257071	0.402515144	-0.104864556
37	38	39	40	41	42
-0.445578216	0.122029823	-1.239434362	0.173722439	-0.169315133	0.084309118
43	44	45	46	47	48
0.634897121	0.045611531	0.281440234	0.262221602	-1.326367935	0.406019732
49	50	51	52	53	54
0.657535525	0.799136326	0.508881711	0.777137269	0.607190479	0.213506462
55	56	57	58	59	60
-1.053890080	0.533794522	0.659021739	1.106966418	0.349255331	0.664164401
61	62	63	64	65	66
1.156981333	-0.186313946	-0.472220598	-0.169799778	0.817697337	0.659232033
67	68	69	70	71	72
0.495900852	0.446994143	1.333581904	0.800196617	0.031396990	0.875994660
73	74	75	76	77	78
0.265074697	-0.275907564	-0.900639088	-0.251363139	0.585489964	0.323482255
79	80	81	82	83	84

```

0.019046625 0.187758777 1.286652644 0.820813657 0.162247460 0.339525077
      85      86      87      88      89      90
0.947249952 -0.970953240 1.640820376 0.778254854 0.007237055 0.726087105
      91      92      93      94      95      96
0.899128837 0.784777498 0.397159326 -1.026992978 1.062752087 1.435200389
      97
0.678537942
> plot(fitted(lm_model), resid(lm_model), main = " Vesinv Scatter plot")
+ "
+
Error: unexpected end of input
> plot(fitted(lm_model), resid(lm_model), main = " Vesinv Scatter plot")
> plot(fitted(lm_model), resid(lm_model), main = " Scatter plot")
> plot(fitted(lm_model), resid(lm_model), main = " Linear Model Scatter plot")
> abline(h=0)
> qqnorm(resid(lm_model), main = " Linear Model Q-Q plot")
> qqline(resid(lm_model))
> #Stepwise Selection with AIC
> step(nullmd, scope= list(lower=~1, upper=~cancervol
+ + as.factor(vesinv) +
+ capspen +gleason), k = 2)
Start: AIC=28.72
log(Cancer_data$psa) ~ 1

              Df Sum of Sq      RSS      AIC
+ cancervol    1    55.164   72.605 -24.0986
+ as.factor(vesinv) 1    40.984   86.785  -6.7944
+ gleason       1    37.122   90.647  -2.5707
+ capspen       1    34.286   93.482   0.4169
<none>                     127.769  28.7246

Step: AIC=-24.1
log(Cancer_data$psa) ~ cancervol

              Df Sum of Sq      RSS      AIC
+ gleason       1     8.247   64.358 -33.794
+ as.factor(vesinv) 1     6.547   66.058 -31.265
<none>                     72.605 -24.099
+ capspen       1     0.967   71.638 -23.400
- cancervol      1    55.164  127.769  28.725

Step: AIC=-33.79
log(Cancer_data$psa) ~ cancervol + gleason

              Df Sum of Sq      RSS      AIC
+ as.factor(vesinv) 1     4.0178  60.340 -38.047
<none>                     64.358 -33.794
+ capspen          1     0.1685  64.190 -32.048
- gleason           1     8.2468  72.605 -24.099
- cancervol         1    26.2887  90.647  -2.571

Step: AIC=-38.05
log(Cancer_data$psa) ~ cancervol + gleason + as.factor(vesinv)

              Df Sum of Sq      RSS      AIC
<none>                     60.340 -38.047
+ capspen          1     0.3013  60.039 -36.532
- as.factor(vesinv) 1     4.0178  64.358 -33.794
- gleason           1     5.7179  66.058 -31.265
- cancervol         1    12.7041  73.044 -21.513

Call:
lm(formula = log(Cancer_data$psa) ~ cancervol + gleason + as.factor(vesinv),
    data = Cancer_data)

Coefficients:
(Intercept)      cancervol      gleason as.factor(vesinv)1
    -0.72120         0.05981         0.38491          0.62117

```

```
> aic_lm_model <- lm(formula = log(Cancer_data$psa) ~ cancervol + gleason +
+ as.factor(vesinv),
+ data = Cancer_data)
> plot(fitted(aic_lm_model), resid(aic_lm_model), main = "AIC Linear Model Scatter plot")
Error: unexpected ')' in "plot(fitted(aic_lm_model), resid(aic_lm_model), main = "AIC Linear Model
Scatter plot")"
> abline(h=0)
> plot(fitted(aic_lm_model), resid(aic_lm_model), main = "AIC Linear Model Scatter plot")
> abline(h=0)
> qqnorm(resid(aic_lm_model) , main = "AIC Linear Model Scatter plot")
> qqline(resid(aic_lm_model))
> qqnorm(resid(aic_lm_model) , main = "AIC Linear Model Q-Q plot")
> qqline(resid(aic_lm_model))
> nullmd = lm(log(Cancer_data$psa)~1, data= Cancer_data)
> step(nullmd, scope=list(lower=~1, upper=~weight+age+benpros+cancervol+
+ gleason+as.factor(vesinv)+capspen))
Start:  AIC=28.72
log(Cancer_data$psa) ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ cancervol	1	55.164	72.605	-24.0986
+ as.factor(vesinv)	1	40.984	86.785	-6.7944
+ gleason	1	37.122	90.647	-2.5707
+ capspen	1	34.286	93.482	0.4169
+ age	1	3.688	124.080	27.8831
+ benpros	1	3.166	124.603	28.2911
<none>			127.769	28.7246
+ weight	1	1.893	125.876	29.2767

```
Step:  AIC=-24.1
log(Cancer_data$psa) ~ cancervol
```

	Df	Sum of Sq	RSS	AIC
+ gleason	1	8.247	64.358	-33.794
+ benpros	1	7.803	64.802	-33.128
+ as.factor(vesinv)	1	6.547	66.058	-31.265
+ age	1	2.662	69.944	-25.721
+ weight	1	1.790	70.815	-24.520
<none>			72.605	-24.099
+ capspen	1	0.967	71.638	-23.400
- cancervol	1	55.164	127.769	28.725

```
Step:  AIC=-33.79
log(Cancer_data$psa) ~ cancervol + gleason
```

	Df	Sum of Sq	RSS	AIC
+ benpros	1	6.2827	58.075	-41.758
+ as.factor(vesinv)	1	4.0178	60.340	-38.047
+ weight	1	2.0334	62.325	-34.908
<none>			64.358	-33.794
+ age	1	0.9611	63.397	-33.253
+ capspen	1	0.1685	64.190	-32.048
- gleason	1	8.2468	72.605	-24.099
- cancervol	1	26.2887	90.647	-2.571

```
Step:  AIC=-41.76
log(Cancer_data$psa) ~ cancervol + gleason + benpros
```

	Df	Sum of Sq	RSS	AIC
+ as.factor(vesinv)	1	4.8466	53.229	-48.211
<none>			58.075	-41.758
+ weight	1	0.4006	57.675	-40.429
+ capspen	1	0.1863	57.889	-40.069
+ age	1	0.0059	58.070	-39.768
- benpros	1	6.2827	64.358	-33.794
- gleason	1	6.7262	64.802	-33.128
- cancervol	1	29.9589	88.034	-3.407

```
Step:  AIC=-48.21
```

```
log(Cancer_data$psa) ~ cancervol + gleason + benpros + as.factor(vesinv)
```

	Df	Sum of Sq	RSS	AIC
<none>			53.229	-48.211
+ capspen	1	0.3923	52.837	-46.928
+ weight	1	0.3306	52.898	-46.815
+ age	1	0.0250	53.204	-46.256
- gleason	1	4.2389	57.468	-42.778
- as.factor(vesinv)	1	4.8466	58.075	-41.758
- benpros	1	7.1115	60.340	-38.047
- cancervol	1	14.7580	67.987	-26.473

Call:

```
lm(formula = log(Cancer_data$psa) ~ cancervol + gleason + benpros +
    as.factor(vesinv), data = Cancer_data)
```

Coefficients:

	(Intercept)	cancervol	gleason	benpros
	-0.65013	0.06488	0.33376	0.09136
as.factor(vesinv)1	0.68421			

```
> fullmd <- lm(formula = log(Cancer_data$psa) ~ cancervol + gleason + benpros +
+ as.factor(vesinv), data = Cancer_data)
> nullmd = lm(log(Cancer_data$psa)~1, data= Cancer_data)
> step(nullmd,scope=list(lower=~1,upper=~weight+age+benpros+cancervol+
+ gleason+as.factor(vesinv)+capspen))
Start: AIC=28.72
log(Cancer_data$psa) ~ 1
```

	Df	Sum of Sq	RSS	AIC
+ cancervol	1	55.164	72.605	-24.0986
+ as.factor(vesinv)	1	40.984	86.785	-6.7944
+ gleason	1	37.122	90.647	-2.5707
+ capspen	1	34.286	93.482	0.4169
+ age	1	3.688	124.080	27.8831
+ benpros	1	3.166	124.603	28.2911
<none>			127.769	28.7246
+ weight	1	1.893	125.876	29.2767

Step: AIC=-24.1

```
log(Cancer_data$psa) ~ cancervol
```

	Df	Sum of Sq	RSS	AIC
+ gleason	1	8.247	64.358	-33.794
+ benpros	1	7.803	64.802	-33.128
+ as.factor(vesinv)	1	6.547	66.058	-31.265
+ age	1	2.662	69.944	-25.721
+ weight	1	1.790	70.815	-24.520
<none>			72.605	-24.099
+ capspen	1	0.967	71.638	-23.400
- cancervol	1	55.164	127.769	28.725

Step: AIC=-33.79

```
log(Cancer_data$psa) ~ cancervol + gleason
```

	Df	Sum of Sq	RSS	AIC
+ benpros	1	6.2827	58.075	-41.758
+ as.factor(vesinv)	1	4.0178	60.340	-38.047
+ weight	1	2.0334	62.325	-34.908
<none>			64.358	-33.794
+ age	1	0.9611	63.397	-33.253
+ capspen	1	0.1685	64.190	-32.048
- gleason	1	8.2468	72.605	-24.099
- cancervol	1	26.2887	90.647	-2.571

Step: AIC=-41.76

```
log(Cancer_data$psa) ~ cancervol + gleason + benpros
```

	Df	Sum of Sq	RSS	AIC
+ as.factor(vesinv)	1	4.8466	53.229	-48.211
<none>			58.075	-41.758
+ weight	1	0.4006	57.675	-40.429
+ capspen	1	0.1863	57.889	-40.069
+ age	1	0.0059	58.070	-39.768
- benpros	1	6.2827	64.358	-33.794
- gleason	1	6.7262	64.802	-33.128
- cancervol	1	29.9589	88.034	-3.407

Step: AIC=-48.21

log(Cancer_data\$psa) ~ cancervol + gleason + benpros + as.factor(vesinv)

	Df	Sum of Sq	RSS	AIC
<none>			53.229	-48.211
+ capspen	1	0.3923	52.837	-46.928
+ weight	1	0.3306	52.898	-46.815
+ age	1	0.0250	53.204	-46.256
- gleason	1	4.2389	57.468	-42.778
- as.factor(vesinv)	1	4.8466	58.075	-41.758
- benpros	1	7.1115	60.340	-38.047
- cancervol	1	14.7580	67.987	-26.473

Call:

```
lm(formula = log(Cancer_data$psa) ~ cancervol + gleason + benpros +
    as.factor(vesinv), data = Cancer_data)
```

Coefficients:

	(Intercept)	cancervol	gleason	benpros
	-0.65013	0.06488	0.33376	0.09136
as.factor(vesinv)1	0.68421			

```
> fullmd <- lm(formula = log(Cancer_data$psa) ~ cancervol + gleason + benpros +
+ as.factor(vesinv), data = Cancer_data)
> plot(fitted(fullmd), resid(fullmd), main = " Final Linear Model Scatter plot")
> abline(h=0)
> qqnorm(resid(fullmd), main = " Final Linear Model Q-Q plot")
> qqline(resid(fullmd))
> pred_cancervol <- mean(cancervol)
> pred_gleason <- mean(gleason)
> pred_benpros <- mean(benpros)
> table(vesinv)
vesinv
 0  1
76 21
> pred_vesinv <- 0
> new <- data.frame(cancervol=pred_cancervol, gleason=pred_gleason, benpros=pred_benpros, vesinv=
pred_vesinv)
> predict(fullmd, newdata = new)
1
2.330541
>
```